

Stochastic kinetics of the circular gene hypothesis: feedback effects and protein fluctuations

R.R. Wadhwa^a, L. Zálányi^b, J. Szenté^c, L. Négyessy^b, P. Érdi^{a,b,*}

^aCenter for Complex Systems Studies, Kalamazoo College, 1200 Academy Street,
Kalamazoo, MI 49006, USA

^bWigner Research Centre for Physics, Hungarian Academy of Sciences, Budapest, Hungary

^cAtmospheric, Oceanic and Space Sciences, College of Engineering, University of Michigan,
Ann Arbor, MI, USA

Abstract

Stochastic kinetic models of genetic expression are able to describe protein fluctuations. A comparative study of the canonical and a feedback model is given here by using stochastic simulation methods. The feedback model is skeleton model implementation of the circular gene hypothesis, which suggests the interaction between the synthesis and degradation of mRNA. Qualitative and quantitative changes in the shape and in the numerical characteristics of the stationary distributions suggest that more combined experimental and theoretical studies should be done to uncover the details of the kinetic mechanisms of gene expressions.

Keywords: genetic expression, stochastic kinetics

2010 MSC: 80A30

1. Introduction

Protein availability is a *condicio sine qua non* of cellular processes and survival, and is determined by gene regulation. Gene regulation contains many biochemical and biophysical processes. While traditional biochemistry adopted a rather rigid deterministic scenario considering the execution of instructions encoded in DNA, chemical reactions taking place at the single cell level are now admittedly better described by stochastic models than by deterministic ones. Reactions in gene expression, such as promoter activity and inactivity, transcription, translation, and decaying of mRNA and proteins are the most important chemical steps. Measurements on stochastic gene expression in single cells with single molecule sensitivity [6, 9] implied the necessity of stochastic description [22]. Since our goal here is to contribute to the understanding of the nature of protein fluctuations, only stochastic kinetic modeling technique can be relevant.

*Corresponding author: +1 (269) 337-5720

Email address: perdi@kzoo.edu (P. Érdi)

The perspective that models of gene expression should have stochastic elements goes back to the pioneering works of D. Rigney and O. Berg [3, 23–25], but these works came too early for mainstream molecular biologists. Stochastic chemical kinetics became the *lingua franca* of modeling gene regulatory networks and related fields twenty years later due to highly cited papers [2, 18].

Stochastic models proved to be very efficient to study the kinetic mechanisms of genetics and, more generally, systems biological processes [10]. Measured fluctuations in reactions two sources [9] (i) **intrinsic** noise is related to variations in protein levels even in a population of cells with identical genotype and concentrations and states of cellular components, (ii) **extrinsic** noise due to fluctuations in the amount or activity of molecules involved in the expression of a gene, like RNA polymerase or ribosomes. The reaction system, what might be called the canonical model of gene expression [20], belongs to the category of compartmental models. Such systems are characterized by the fact that the activity of one molecular entity is independent of the other entities. In other words, no interaction between any two such entities occurs. Such models can fully be solved (i.e. the time-dependent moments can be calculated) by using the generating function method. More specifically, the different sources of protein fluctuations have been calculated [20]. Under not too restrictive conditions it was found [6, 11] that the stationary distribution of the protein fluctuation can be well approximated by gamma distribution. However, as gamma distribution is a very general one, alterations of the reactions system and/or the rate constants may imply changes in the shape and parameters of the stationary distributions.

Realistic models should take into account feedback, burst, delay, etc. mechanisms too, and some exact results are available for specific families of models [4, 11, 14, 21, 27]. These models contain bimolecular reaction steps too, so the compartmental kinetic framework based on independent activities cannot be assumed anymore. Linear noise approximation is often used to calculate protein fluctuations, but its reliability for systems containing bimolecular reaction steps is restricted [29].

Simulation methods (for a recent short review see 2.6 of [10]) are appropriate tools to obtain information about the size and nature of fluctuations as they help overcome the limitations of the methods described above.

A recent conceptually new hypothesis [13] suggested that gene expression might be circular, since the degradation and synthesis of mRNA seem to be interconnected by a feedback mechanism. Due to the lack of available kinetic data the hypothesis cannot be falsified for the time being. However, as the metabolism of mRNA is better described by some bimolecular reactions, it might affect protein fluctuations. Specifically, by setting a secondary mechanism to promote mRNA synthesis may increase the lifetime ratio of the lifetimes of mRNA and of proteins, which increases protein fluctuation. Our question was whether or not the feedback mechanism has a significant effect on protein fluctuations. If yes, it is worth studying the details.

2. Biological background

Gene expression is the complicated process of converting genetic information from a DNA sequence into proteins. In eukaryotes, DNA is located in the cell nucleus. Prokaryotes do not have a nucleus, and DNA can be found in the cytoplasm. In prokaryotes there are two main processes in gene expression: **transcription** and **translation**. In eukaryotes, there is an additional process: **splicing**.

Transcription is a series of events that use DNA to synthesize messenger RNA (mRNA) by using the enzyme RNA polymerase as a catalyst. The series of events contain (in prokaryotes) *binding*, *initiation*, *RNA synthesis*, *elongation*, and *termination*. Specifically, a **promoter** is a region of DNA where binding of transcription factor proteins initiate transcription. Eukaryotic transcription is much more complicated, but depends on these basic steps.

Splicing is a modification of the nascent mRNA transcript in which certain nucleotide sequences (*introns*) are removed while other sequences (*exons*) remain.

Translation is a process in which the is read out from the mRNA by the ribosome complex and translated into the amino acid sequence in proteins (with the help of tRNA). It contains more elementary steps, such as *initiation*, *elongation*, *translocation*, and *termination*.

Degradation: although DNA is stable, RNA and protein molecules are subject to degradation. It is an important step in the regulation of gene expression and fluctuation of protein concentration.

A recent hypothesis [13] suggested that eukaryotic gene expression can be viewed as a circular process, where transcription and mRNA degradation are interconnected. The big question is, “How could mRNA synthesis in the nucleus and mRNA decay in the cytoplasm be mechanistically linked?” [5]. Possible mechanisms of coupling mRNA synthesis and decay have been analyzed [5, 19]. The 5' to 3' exoribonuclease *xrn1*, a large protein involved in cytoplasmatic mRNA degradations might be a critical component [19], and it may play a dual role in some subprocesses of transcription, namely in initiation and elongation.

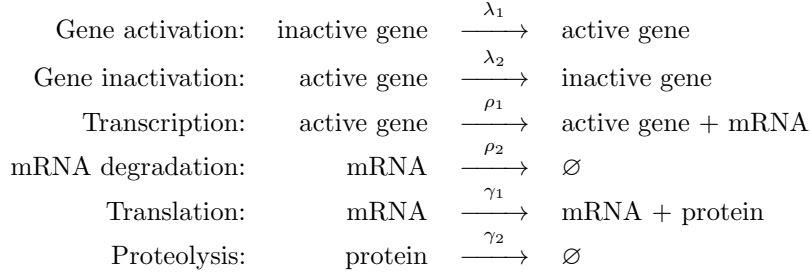
Based on these observations about the dual role of *xrn1* in transcription [13, 19], a minimal model that takes into account feedback effects has been set by including three more steps: promoter assignment, promoter reassignment, and *xrn1* dependent transcription.

In this paper, the nature of protein fluctuation in the canonical model and a simple feedback model implementing the dual role of *xrn1* is studied using stochastic simulations. Based on the results, we predict that the feedback process has significant effects on the fluctuations by the additive effects of the enhanced mRNA fluctuations, so the detailed mechanisms should be studied by combined experimental and modeling studies.

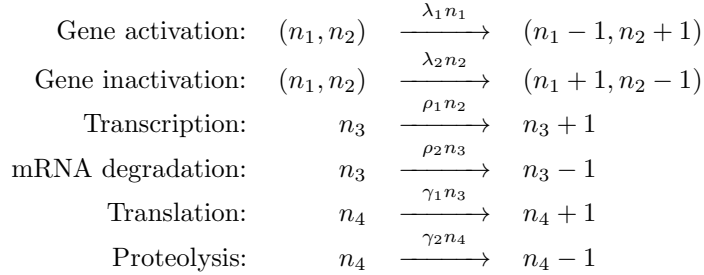
3. The model

3.1. Canonical model

Gene expression can be modeled as a three-stage process: gene activation, transcription, and translation. These are coupled by the opposite processes of gene inactivation, mRNA degradation, and proteolysis, respectively. Gene expression can be modeled as a reaction system of three chemical species, slightly modified from the existing schematic for the canonical model [4].



In this system, all the reactions are first order. This reaction system can be alternatively defined using the number of each chemical species present in a cell. Here n_1 , n_2 , n_3 , and n_4 represent the number of inactive genes, active genes, mRNAs, and proteins in the cell, respectively [20].



To determine the value of the rate constants, we refer to experimental results, using *E. coli* as our model organism. The half-life of mRNA in *E. coli*, calculated as the natural logarithm of 2 divided by the rate constant for mRNA degradation, is between 3 min and 8 min [4]. Using a timescale measured in seconds, we choose the average half-life of an mRNA molecule in the simulation to be 300 s. This leads to a value of $\ln(2)/300 \approx 0.00231$ for ρ_2 .

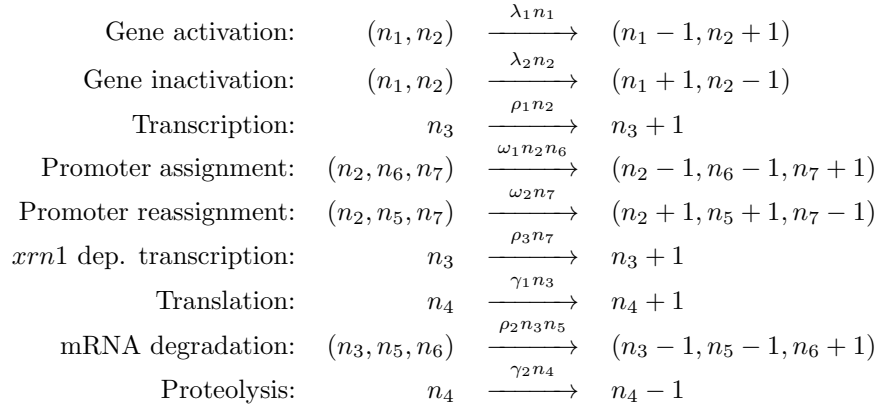
There are indications from experimental data that ρ_1/ρ_2 is variable in *E. coli*, ranging from 1 to a few dozen - we assume a value of 10 for this ratio, which implies $\rho_1 \approx 0.0231$ [4, 16]. It has also been experimentally determined that for proteins in *E. coli*, the average value of γ_1/γ_2 is 540 [4, 16]. For this simulation, we choose $\gamma_1 = 0.14$ as it approaches a stationary state with sufficient speed. For the sake of simplicity, we assume there exists only a single copy of the gene we are interested in. We also choose $\lambda_1 = 1$ and $\lambda_2 = 7$, although we shall see that the choice of values for λ_1 and λ_2 is arbitrary. The ratio $\lambda_1/(\lambda_1 + \lambda_2)$

indicates the proportion of time for which a gene is active [20], and is what truly matters.

The initial value of (n_1, n_2, n_3, n_4) for the reaction system was $(1, 0, 0, 0)$.

3.2. Feedback model: the role of *xrn1*

Gene expression involving *xrn1* requires a model with more reactions to be accurately modeled. Using the biological background given in Section 2, we give the following model to account for feedback due to *xrn1*. Here n_5 , n_6 , and n_7 represent the number of *xrn1* molecules, *xrn1* complexes, and *xrn1* binding to the promoter in the cell, respectively.



It is experimentally supported that the rate constant of *xrn1* dependent transcription is equal to the rate constant of transcription, implying that $\rho_3 = \rho_1 = 0.0231$ [13]. The effects of varying values of ω_1 and ω_2 on the resulting protein distribution are investigated in Section 4. The remaining rate constants in the feedback model have values identical to the corresponding rate constants in the canonical model.

The initial value of $(n_1, n_2, n_3, n_4, n_5, n_6, n_7)$ for the reaction system was $(1, 0, 0, 0, 10, 0, 0)$.

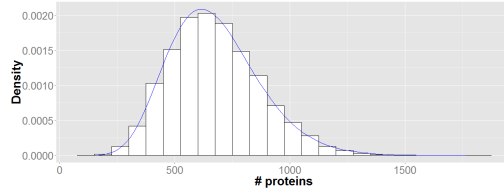


Figure 1: The distribution of proteins in the canonical model at $t = 100,000$ s with $n = 10,000$ stochastic trajectories. The plotted line is the best-fit gamma distribution with a shape parameter of 11.59 ± 0.16 and a rate parameter of 0.0172 ± 0.0002 . The expected value is equal to 674.54 and the standard deviation of the distribution is 196.80.

Rate Constant	Value
λ_1	1.0
λ_2	7.0
ρ_1	0.0231
ρ_2	$\rho_1/10$
γ_1	0.14
γ_2	$\gamma_1/540$

Table 1: The rate constants used to produce the protein distribution of the canonical model on the left. ρ_1 has been experimentally determined, as have the ratios ρ_1/ρ_2 and γ_1/γ_2 [4, 16].

3.3. Simulation method

All simulations for this paper were conducted in the Cain interface developed by Sean Mauch [17]. Realizations of the stochastic process were produced using Gillespie’s direct method [12]. Considering a reaction system as a set of ordinary differential equations as defined by the laws of mass action kinetics, numerical integration using the Cash-Karp variant of the Runge-Kutta method was conducted to produce deterministic trajectories [7]. All histograms and plots were produced using the ggplot2 package in the R programming language with the multiplot function from the Cookbook for R website [1, 30]. Function `fitdistr()` in the MASS package from the R programming language was used to find the best fit parameters for the distributions.

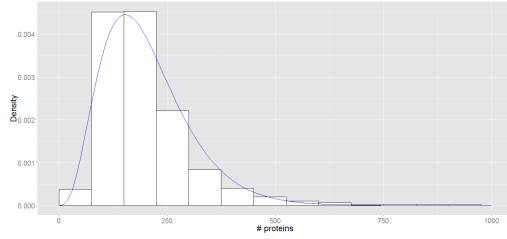


Figure 2: The distribution of proteins in the feedback model at $t = 100,000$ s with $n = 10,000$ stochastic trajectories. Plotted lines are the best-fit gamma distribution with a shape parameter of 4.089 ± 0.06 and a scale parameter of 0.020 ± 0.0003 . The expected value is equal to 203 and the standard deviation of the distribution is 124.29.

Rate Constant	Value
λ_1	1.0
λ_2	7.0
ρ_1	0.0231
ρ_2	$\rho_1/10$
ρ_3	ρ_1
ω_1	1
ω_2	$3\rho_1/2$
γ_1	0.14
γ_2	$\gamma_1/540$

Table 2: The rate constants used to produce the protein distribution of the feedback model on the left. ρ_1 and ρ_3 have been experimentally determined, as have the ratios ρ_1/ρ_2 and γ_1/γ_2 [4, 16].

4. Simulation Results

Stationary protein distributions for both models were plotted as histograms, and fitted with gamma distributions. Figure 1 shows the steady state protein distribution for the canonical model with the typical parameter values given in Table 1. Figure 2 shows the steady state distribution for the feedback model with the typical parameter values given in Table 2. The increased skew in the steady state protein distribution of the feedback model as compared to the canonical model is observed in all the steady state distributions explored.

4.1. Varying the rate of gene (in)activation: λ_1 and λ_2

As stated in Paulsson (2005), the kinetic dynamics of gene activation and inactivation can be approximately modeled as a random telegraph process, where each gene independently switches on with rate λ_1 and switches off with rate λ_2 . The probability of a gene being on is given by $P_{\text{on}} = \lambda_1/(\lambda_1 + \lambda_2)$. Correspondingly, the expected number of proteins in steady state should be proportional to P_{on} . This pattern is confirmed by Figure 3, wherein the expected number of

	1/16	1/8	1/4	1/2
0.01	337.5	675	1350	2700
0.05	337.5	675	1350	2700
0.14	337.5	675	1350	2700
0.50	337.5	675	1350	2700

Table 3: Expected number of proteins as given by deterministic solutions of the canonical model. The horizontal labels show varying values of $\lambda_1/(\lambda_1 + \lambda_2)$. The vertical labels show varying values of γ_1 . $\rho_1 = 0.0231$; $\rho_2 = \rho_1/10$; $\gamma_2 = \gamma_1/540$.

	1/16	1/8	1/4	1/2
0.01	284.3 \pm 41.0	568.2 \pm 58.2	1136.8 \pm 83.1	2275.1 \pm 116.5
0.05	336.8 \pm 84.8	675.6 \pm 120.5	1350.7 \pm 173.9	2702.0 \pm 241.1
0.14	338.2 \pm 139.6	676.4 \pm 194.7	1349.7 \pm 274.6	2701.0 \pm 387.3
0.50	335.1 \pm 223.1	675.9 \pm 326.2	1342.3 \pm 455.3	2705.2 \pm 642.1

Table 4: Expected number of proteins \pm standard deviation of the steady state protein distribution as given by stochastic simulations of the canonical model. The horizontal labels show varying values of $\lambda_1/(\lambda_1 + \lambda_2)$. The vertical labels show varying values of γ_1 . $\rho_1 = 0.0231$; $\rho_2 = \rho_1/10$; $\gamma_2 = \gamma_1/540$.

proteins at steady state for the deterministic and stochastic results is linearly proportional to P_{on} . Due to the presence of the relationship described by Paulson and confirmation in the canonical model, we arbitrarily choose the values $\lambda_1=1$ and $\lambda_2=15$ when exploring the more complicated feedback model. Results for other values of λ_1 and λ_2 can be extrapolated by finding the value of P_{on} for the scenario in question.

4.2. Varying the translation rate: γ_1

Interestingly, the deterministic solutions of the canonical and feedback models are independent of γ_1 (Figure 3A). In the stochastic results of the canonical model, the number of proteins at steady state is lower at $\gamma_1 = 0.01$; for values of $\gamma_1 = 0.05, 0.14, 0.50$ the stochastic results seem to converge to the deterministic solution (Figure 3B). However, as seen in Table 4, despite this convergence, the standard deviation of the steady state protein distribution increases as γ_1 increases. These results suggest that using a single value for the number of proteins at steady state as given by the deterministic solution becomes an increasingly worse approximation of the true distribution of proteins in a cell as the translation rate increases.

The convergence pattern does not hold true of the feedback model. Figure 4 shows an instance where the deterministic and stochastic solutions clearly diverge irrespective of the value of γ_1 . Thus, for the feedback model, the deterministic and stochastic methods give distinct values for the expected number of proteins at steady state.

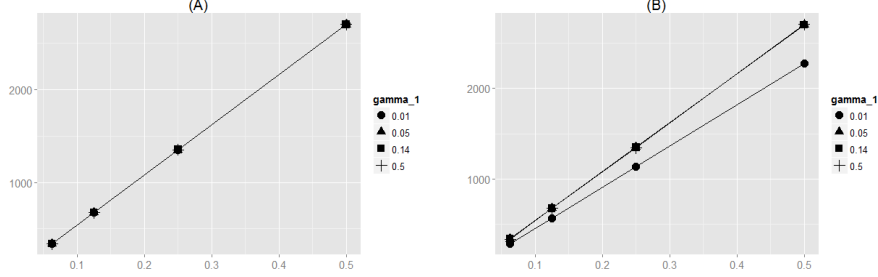


Figure 3: The x-axis on both plots is $\lambda_1/(\lambda_1+\lambda_2)$; the y-axis on both plots is the mean number of proteins at steady state. (A) Deterministic solutions for the canonical model. The plotted series all match exactly (Table 3). (B) Stochastic simulation results for the canonical model. The series for $\gamma_1 = 0.05, 0.14$, and 0.50 closely overlap (Table 4).

4.3. Varying the rate of promoter assignment: ω_1

Based on chemical reaction representing promoter assignment, increasing the value of ω_1 decreases the number of active genes by 1, decreases the number of *xrn1* complexes by 1, and increasing the number of *xrn1* binding to promoters by 1. Decreasing the number of active genes results in reduced occurrence of transcription; decreasing the number of *xrn1* complexes reduces promoter assignment in a feedback loop; increasing *xrn1* binding to promoters increases *xrn1*-based transcription. Thus, the first effect (which does not directly involve *xrn1*) should decrease the number of proteins at steady state, while the last effect (which directly involves *xrn1*) should increase the number of proteins at steady state. Overall, as ω_1 increases, the expected number of proteins at steady state decreases. This indicates that changing the number of active genes by a fixed amount has a greater effect on the number of proteins at steady state than does changing any of the *xrn1* elements by the same fixed amount. In essence, although adding *xrn1*-based processes to the canonical model results in an effect, the basic processes present in both models still have the majority effect in determining steady state protein levels. Most of the change in protein levels happens between $\omega_1 = 0.10$ and 0.50 .

In addition to affecting the mean of steady state protein distribution, increasing ω_1 also results in an overall decrease in the standard deviation of the steady state protein distribution. However, for larger values of γ_1 (e.g. $\gamma_1 = 0.50$, Table 9) the standard deviation is either approximately equal to or larger than the expected number of proteins at steady state. This indicates a high degree of variation in the steady state distribution. Therefore, even though exploring the effect of varying ω_1 in the deterministic solution (Figure 5E) shows the same trend as it does in the stochastic results, due to the relatively large values of the standard deviation, failing to account for variability in the data is a significant con of the deterministic solution.

	$1.25\rho_1$	$1.50\rho_1$	$1.75\rho_1$	$2.00\rho_1$
0.1	156.2	97.9	77.5	67.0
0.5	141.1	92.3	74.1	64.5
1.0	139.4	91.7	73.7	64.2
1.5	138.8	91.4	73.5	64.1

Table 5: Expected number of proteins as given by deterministic solutions of the feedback model. The horizontal labels show varying values of ω_2 . The vertical labels show varying values of ω_1 . Much like the canonical model (Table 3), deterministic solutions of the feedback model are independent of the value of γ_1 . $\gamma_1 = 0.01, 0.05, 0.14$, and 0.50 resulted in the exact same values for expected number of proteins.

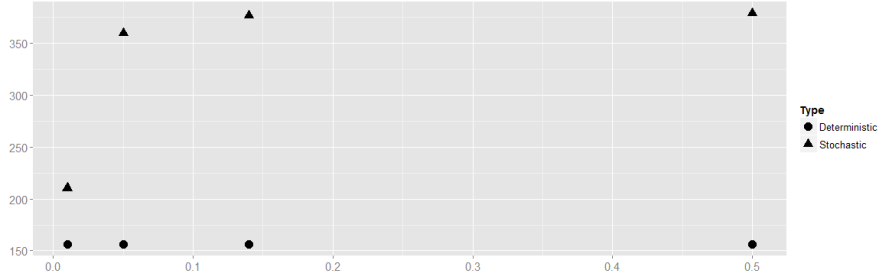


Figure 4: Deterministic and stochastic solutions for number of proteins at steady state with the given parameter values: $\lambda_1 = 1$; $\lambda_2 = 15$; $\omega_1 = 0.01$; $\omega_2 = 1.25\rho_1$. The x-axis on both plots is γ_1 ; the y-axis on both plots is the mean number of proteins at steady state.

	$1.25\rho_1$	$1.50\rho_1$	$1.75\rho_1$	$2.00\rho_1$
0.1	210.3 ± 230.9	78.3 ± 40.8	53.3 ± 18.2	44.0 ± 12.8
0.5	155.2 ± 135.0	68.0 ± 27.9	49.5 ± 15.3	41.5 ± 11.6
1.0	150.2 ± 131.4	66.8 ± 26.7	48.6 ± 14.7	41.0 ± 11.5
1.5	147.8 ± 125.9	65.8 ± 25.3	48.5 ± 14.5	40.9 ± 11.2

Table 6: Expected number of proteins \pm standard deviation of the steady state protein distribution as given by stochastic simulations of the feedback model. The horizontal labels show varying values of ω_2 . The vertical labels show varying values of ω_1 . $\lambda_1/(\lambda_1 + \lambda_2) = 1/15$; $\rho_1 = 0.0231$; $\rho_2 = \rho_1/10$; $\rho_3 = \rho_1$; $\gamma_1 = 0.01$; $\gamma_2 = \gamma_1/540$.

	$1.25\rho_1$	$1.50\rho_1$	$1.75\rho_1$	$2.00\rho_1$
0.1	359.9 ± 482.6	128.0 ± 87.9	88.3 ± 40.5	72.4 ± 27.3
0.5	261.6 ± 307.5	112.8 ± 62.0	81.9 ± 35.1	68.6 ± 24.6
1.0	249.5 ± 288.7	110.3 ± 63.0	79.9 ± 32.4	67.5 ± 23.9
1.5	248.8 ± 310.0	108.7 ± 58.3	80.1 ± 32.4	67.3 ± 23.7

Table 7: Expected number of proteins \pm standard deviation of the steady state protein distribution as given by stochastic simulations of the feedback model. The horizontal labels show varying values of ω_2 . The vertical labels show varying values of ω_1 . $\lambda_1/(\lambda_1 + \lambda_2) = 1/15$; $\rho_1 = 0.0231$; $\rho_2 = \rho_1/10$; $\rho_3 = \rho_1$; $\gamma_1 = 0.05$; $\gamma_2 = \gamma_1/540$.

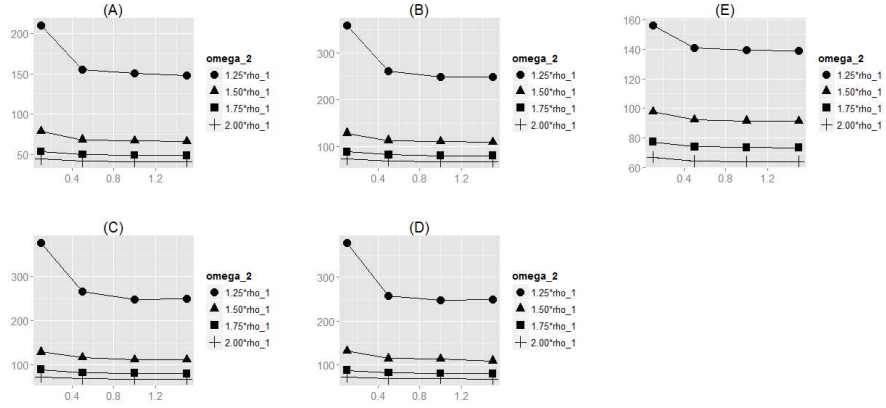


Figure 5: Deterministic (E) and stochastic (A-D) solutions for number of proteins at steady state with the given parameter values: $\lambda_1 = 1$; $\lambda_2 = 15$; $\rho_1 = 0.0231$. The x-axis on all plots is ω_1 ; the y-axis on all plots is the mean number of proteins at steady state. (A) $\gamma_1 = 0.01$; (B) $\gamma_1 = 0.05$; (C) $\gamma_1 = 0.14$; (D) $\gamma_1 = 0.50$; (E) Deterministic solution is independent of the value of γ_1 .

	$1.25\rho_1$	$1.50\rho_1$	$1.75\rho_1$	$2.00\rho_1$
0.1	376.6 ± 741.3	129.4 ± 137.1	89.3 ± 64.5	72.8 ± 44.1
0.5	265.7 ± 449.3	116.5 ± 107.2	82.2 ± 52.8	69.0 ± 38.4
1.0	247.7 ± 398.9	111.9 ± 98.8	81.3 ± 52.2	68.2 ± 36.9
1.5	249.0 ± 407.0	111.6 ± 98.9	80.4 ± 49.7	68.0 ± 38.9

Table 8: Expected number of proteins \pm standard deviation of the steady state protein distribution as given by stochastic simulations of the feedback model. The horizontal labels show varying values of ω_2 . The vertical labels show varying values of ω_1 . $\lambda_1/(\lambda_1 + \lambda_2) = 1/15$; $\rho_1 = 0.0231$; $\rho_2 = \rho_1/10$; $\rho_3 = \rho_1$; $\gamma_1 = 0.14$; $\gamma_2 = \gamma_1/540$.

	$1.25\rho_1$	$1.50\rho_1$	$1.75\rho_1$	$2.00\rho_1$
0.1	379.1 ± 903.7	131.6 ± 209.7	87.7 ± 101.2	72.9 ± 72.5
0.5	257.1 ± 597.4	114.5 ± 164.4	81.9 ± 90.6	68.3 ± 65.8
1.0	248.5 ± 535.6	113.4 ± 161.9	80.7 ± 92.8	69.0 ± 67.2
1.5	249.0 ± 577.7	109.2 ± 156.8	79.6 ± 85.7	67.5 ± 65.8

Table 9: Expected number of proteins \pm standard deviation of the steady state protein distribution as given by stochastic simulations of the feedback model. The horizontal labels show varying values of ω_2 . The vertical labels show varying values of ω_1 . $\lambda_1/(\lambda_1 + \lambda_2) = 1/15$; $\rho_1 = 0.0231$; $\rho_2 = \rho_1/10$; $\rho_3 = \rho_1$; $\gamma_1 = 0.50$; $\gamma_2 = \gamma_1/540$.

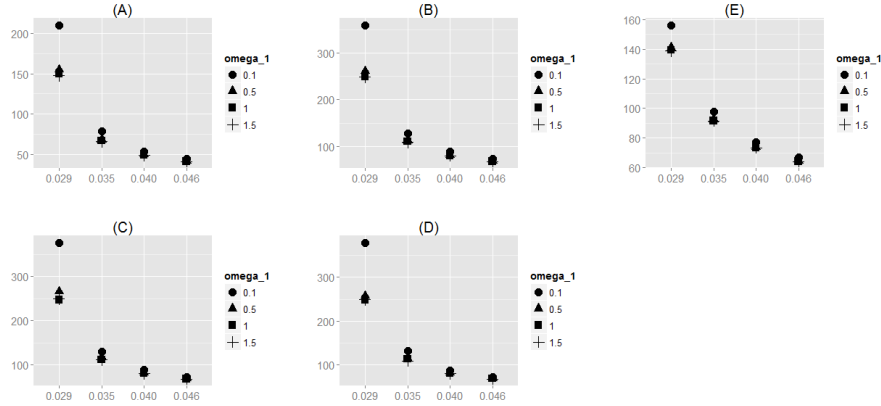


Figure 6: Deterministic (E) and stochastic (A-D) solutions for number of proteins at steady state with the given parameter values: $\lambda_1 = 1$; $\lambda_2 = 15$; $\rho_1 = 0.0231$. The x-axis on all plots is ω_2 ; the y-axis on all plots is the mean number of proteins at steady state. (A) $\gamma_1 = 0.01$; (B) $\gamma_1 = 0.05$; (C) $\gamma_1 = 0.14$; (D) $\gamma_1 = 0.50$; (E) Deterministic solution is independent of the value of γ_1 .

4.4. Varying the rate of promoter reassignment: ω_2

Varying ω_2 has parallel qualitative results to varying ω_1 . Increasing the value of ω_2 decreases the expected number of proteins at steady state, albeit at a different rate. Additionally, the largest change in the expected number of proteins occurs at lower values of ω_2 (between $\omega_2 = 1.25\rho_1$ and $1.50\rho_1$). Like ω_1 , increasing ω_2 reduces the standard deviation of the steady state protein distribution (Tables 6, 7, 8, 9), yet the standard deviation remains large enough relative to the expected value to merit the use of stochastic results over the deterministic solution. However, the deterministic solution does preserve the general trend observed in the stochastic results (Figure 6).

5. Discussions

Stochastic chemical kinetics now has a renaissance due to the consequence of the emergence and development of systems biology. It looks to be one of the most important modeling tool to understand and describe the mechanism of gene expression. While it is one of the basic processes of life, we are far from having a detailed kinetic mechanism of the whole process composed of many subprocesses. Generally a kinetic mechanism is said to be “known”, if all elementary reactions and their rate constants are determined.

Genetic expression is modeled by lumped kinetic models. In a *lumped* model, one step contains a sequence of more elementary reaction steps. The canonical model of genetic expression [20] is technically a compartmental system, and its stochastic model can be completely solved. However, the incorporation of other steps of course implies changes in the kinetic properties of the system under investigation. More specifically, as it was stated recently “protein distribution shape informs on molecular mechanism” [28]. By following the same logic, we were interested in the qualitative (modality, skewness etc.) and quantitative features of the stationary distributions of different models.

A comparative analysis of the canonical and a feedback model was given here. The construction of the feedback model has been motivated by the circular gene expression hypothesis [13], which assumes a mechanism of the interaction between the degradation and synthesis of mRNA. In the model we incorporated three lumped reactions, such as promoter assignment, promoter reassignment, and a second transcription step, which depend on the large protein *xrn1*. The collection and estimation of rate constants is not easy. The data used here is based on *E. coli* as a model organism, and the results could be different for eukaryotes and other organisms. There are initial encouraging results for obtaining more quantitative data [8, 15, 26] and there is a hope that it will be possible to give more reliable and consistent estimation of the rate constants. As concerns the analysis of the model, we restricted ourselves here for simulation studies and for the analysis of these results. Stationary distributions have been empirically constructed from the set of the individual realizations.

How to interpret the results? While our main goal was to see the whether there are characteristic differences between the canonical and the feedback models, remarkable effects of the some changes in the rate constants were also

observed. Most interestingly, the increase of translational rate in the canonical model destroys gamma distribution and leads to the emergence of some kinds of multimodality. It is important to note that the corresponding deterministic model leads to uni-stationarity (and not multi-stationarity). As the realizations show the transient behavior, a system is generally in one of the two possible “high” and “low” states with rapid jumps between them. In the canonical model we don’t see multimodality, but exponential distribution was fitted well. Increased transcription rate implies more expressed right-skewness, in both model, while increased values of the promoter reassignment rate result in a decreased expected value of the protein distribution. The systematic exploration of the three-dimensional parameter space of the rates of the additional reactions of the feedback model should be the next step.

In summary, our studies support the view that qualitative and quantitative changes in the shape and in the numerical characteristics of the stationary distributions of the stochastic models occur due to the consequence of altered reaction network and rate constants. Combined experimental and theoretical studies could help to uncover the details of the kinetic mechanism of the circular gene hypothesis.

Acknowledgements

Thanks for Mordechai (Motti) Choder for motivation and initial correspondence. PE thanks to the Henry Luce Foundation to let him to serve as a Henry R Luce Professor.

References

- [1] Cookbook for r, <http://www.cookbook-r.com/>, accessed: 2014-12-14.
- [2] A. Arkin, J. Ross, H. H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells, *Genetics* 149 (1998) 1633–48.
- [3] O. Berg, A model for the statistical fluctuations of protein numbers in a microbial population, *J. Theor. Biol.* 71 (1978) 587–603.
- [4] P. Bokes, J. R. King, A. T. Wood, M. Loose, Exact and approximate distributions of protein and mrna levels in the low-copy regime of gene expression, *J. Math. Biol.* 64 (2012) 829–854.
- [5] K. A. Braun, E. T. Young, Coupling mrna synthesis and decay, *Mol. Cell. Biol.* 34 (2014) 4078–87.
- [6] L. Cai, N. Friedman, X. S. Xie, Stochastic protein expression in individual cells at the single molecule level, *Nature* 440 (2006) 358–362.
- [7] J. R. Cash, A. H. Karp, A variable order runge-kutta method for initial value problems with rapidly varying right-hand sides, *ACM Trans. Math. Softw.* 16 (1990) 201–222.
- [8] K. Choudhary, S. Oehler, A. Narang, Protein distributions from a stochastic model of the lac operon of e. coli with dna looping: Analytical solution and comparison with experiments, *PLoS ONE* 9 (2014) e102580.
- [9] M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell, *Science* 297 (2002) 1183–1186.
- [10] P. Érdi, G. Lente, *Stochastic Chemical Kinetics. Theory and (Mostly) Systems Biological Applications*, Springer Series in Synergistics, Springer, 2014.
- [11] N. Friedman, L. Cai, S. Xie, Linking stochastic dynamics to population distribution: An analytical framework of gene expression, *Phys. Rev. Lett.* 97 (2006) 168302.
- [12] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.* 81 (1977) 2340–2361.
- [13] G. Haimovich, D. A. Medina, S. Z. Causse, M. Garber, G. Millan-Zambrano, O. Barkai, S. Chavez, J. E. Perez-Ortin, X. Darzacq, M. Choder, Gene expression is circular: Factors for mrna degradation also foster mrna synthesis, *Cell* 153 (2013) 1000–1011.
- [14] N. Kumar, T. Platini, R. Kulkarni, Exact distributions for stochastic gene expression models with bursting and feedback, *arXiv:1409.3499v2*.

- [15] W. Lee, W. Tzou, Computational methods for discovering gene networks from expression data, *Brief Bioinform.* 10 (2009) 408–423.
- [16] P. Lu, C. Vogel, R. Wang, X. Yao, E. M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation, *Nat. Biotechnol.* 25 (2007) 117–124.
- [17] S. Mauch, M. Stalzer, Efficient formulations for exact stochastic simulation of chemical systems, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 8 (2011) 27–35.
- [18] H. H. McAdams, A. Arkin, Stochastic mechanisms in gene expression, *Proc. Natl. Acad. Sci. U. S. A.* 94 (1997) 814–819.
- [19] D. Medina, A. Jordn-Pla, G. Millán-Zambrano, S. Chávez, M. Choder, J. Pérez-Ortn, Cytoplasmic 5 – 3 exonuclease *xrn1p* is also a genome-wide transcription factor in yeast, *Front Genet.* 5 (2014) 1.
- [20] J. Paulsson, Models of stochastic gene expression, *Phys. Life Rev.* 2 (2005) 157–175.
- [21] H. Pendar, T. Platini, R. V. Kulkarni, Exact protein distributions for stochastic models of gene expression using partitioning of poisson processes, *arXiv:1304.6570v1*.
- [22] H. Qian, Stochastic physics, complex systems and biology, *Quant. Biol.* 1 (2013) 50–53.
- [23] D. Rigney, Note on the kinetics and stochasticity of induced protein synthesis as influenced by various models for messenger rna degradation, *J. Theor. Biol.* 79 (1979) 247–257.
- [24] D. Rigney, Stochastic model of constitutive protein levels in growing and dividing bacterial cells, *J. Theor. Biol.* 76 (1979) 453–480.
- [25] D. Rigney, W. Schieve, Stochastic model of linear, continuous protein-synthesis in bacterial populations., *J. Theor. Biol.* 69 (1977) 761–766.
- [26] B. Schwanhuser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, C. W, M. Selbach, Global quantification of mammalian gene expression control, *Nature* 473 (2011) 337342.
- [27] V. Shahrezaei, P. Swain, Analytical distributions for stochastic gene expression, *Proc Natl Acad Sci U S A* 105 (2008) 17256–61.
- [28] M. Sherman, B. Cohen, A computational framework for analyzing stochasticity in gene expression, *PLoS Comput Biol.* 10 (2014) e1003596.
- [29] P. Thomas, H. Matuschek, R. Grima, How reliable is the linear noise approximation of gene regulatory networks?, *BMC Genomics* (2013) Suppl 4:S5.

- [30] H. Wickham, ggplot2: elegant graphics for data analysis, Springer New York, 2009.
URL <http://had.co.nz/ggplot2/book>